

---

Citation:

Martin, A and Griffiths, AW and Meads, D and Surr, C (2019) How should we capture health state utility in dementia? Comparisons of DEMQOL-Proxy-U and of self- and proxy-completed EQ-5D-5L. *Value in Health*, 22 (12). pp. 1417-1426. DOI: <https://doi.org/10.1016/j.jval.2019.07.002>

Link to Leeds Beckett Repository record:

<https://eprints.leedsbeckett.ac.uk/id/eprint/6095/>

Document Version:

Article (Accepted Version)

---

The aim of the Leeds Beckett Repository is to provide open access to our research, as required by funder policies and permitted by publishers and copyright law.

The Leeds Beckett repository holds a wide range of publications, each of which has been checked for copyright and the relevant embargo period has been applied by the Research Services team.

We operate on a standard take-down policy. If you are the author or publisher of an output and you would like it removed from the repository, please [contact us](#) and we will investigate on a case-by-case basis.

Each thesis in the repository has been cleared where necessary by the author for third party copyright. If you would like a thesis to be removed from the repository or believe there is an issue with copyright, please contact us on [openaccess@leedsbeckett.ac.uk](mailto:openaccess@leedsbeckett.ac.uk) and we will investigate on a case-by-case basis.

## TITLE

How should we capture health state utility in dementia? Comparisons of DEMQOL-Proxy-U and of self- and proxy-completed EQ-5D-5L

Adam Martin, David Meads, Alys Wyn Griffiths, Claire Surr

## ABSTRACT

### Objective

Investigate the construct validity, criterion validity and responsiveness of DEMQOL-Proxy-U and of self- and proxy-completed EQ-5D-5L.

### Methods

The analysis used a three-wave, individual-level dataset of 1,004 people living with dementia in residential care that included self-completed EQ-5D-5L and formal- and informal-carer proxy-completed EQ-5D-5L and DEMQOL-Proxy-U utility values, as well as other non-utility cognitive (FAST, CDR, CMAI) and HRQOL measures (QOL-AD-NH, QUALID). Construct validity, criterion validity and responsiveness was assessed using correlation, Bland-Altman plots and panel data regression models.

### Results

Self-completed EQ-5D-5L failed to reflect clinically important differences and changes in FAST, CDR and CMAI, but did capture the resident's own view of HRQOL (QOL-AD-NH). As dementia severity increased, collection of EQ-5D-5L-proxy and DEMQOL-Proxy-U data was more feasible than self-completed EQ-5D-5L. These formal- and informal-carer proxy-measures also better reflected changes in FAST, CDR and CMAI but didn't capture the resident's own view of HRQOL (QOL-AD-NH), despite adequately capturing the proxies own view of the resident's HRQOL (QUALID). This indicates discrepancies between a proxies-view and residents-view of the impact tangible declines in health, cognition or functional abilities have on HRQOL. The EQ-5D-5L-proxy and DEMQOL-Proxy-U were generally poor substitutes. Regardless of which proxy completed it, EQ5D-5L-proxy was typically more responsive than DEMQOL-Proxy-U to changes in CDR, FAST and CMAI, indicating that use of DEMQOL-Proxy-U is not always justified.

### Conclusion

Disparities in the measurement properties of different utility measures mean that choices about how to measure utility in trials could affect economic evaluation outcomes and hence how resources are allocated for dementia care.

## CONCISE DESCRIPTION

The measurement properties, validity and responsiveness of self-completed EQ-5D-5L and formal- and informal-carer proxy-completed EQ-5D-5L and DEMQOL-Proxy-U are compared for 1,004 care-home residents with dementia

## HIGHLIGHTS

Please identify 2-3 “Highlights” that illustrate the paper’s contribution to the field. These bulleted statements should address: 1. What is already known about the topic? 2. What does the paper add to existing knowledge? 3. What insights does the paper provide for informing health care-related decision making?

- **Dementia-specific and proxy-completed PBMs have been proposed for use in intervention studies involving people living in residential care, in instances where generic, self-reported PBMs have been deemed inappropriate.**
- **This is the first study to compare their performance using individual-level panel data and the first to examine their responsiveness to changes in five non- preference-based cognitive or quality of life measures.**
- **This study can guide choices about how to measure utility which, given disparities in the performance of different approaches, could critically affect economic evaluation outcomes and hence how resources are allocated for the care of people living with dementia.**

## MAIN PAPER

### 1. Introduction

The NICE Technology Assessment reference case states that health-related quality of life (HRQOL) should be measured using self-completed EQ-5D, which has been shown to be valid, reliable and responsive across numerous disease areas.<sup>(1)</sup> However, NICE may consider other preference-based measures (PBMs) to be informative, including when the self-completed EQ-5D has weak construct validity within a particular patient population. In these cases, proxy-completed or condition-specific utility measures may be used as supplementary evidence.

For people with dementia, proxy-completed PBMs may be justified due to problems with communication recall, time perception and insight.<sup>(2)</sup> One study showed almost half (48%) of respondents with dementia self-reported having no problems on any EQ-5D dimension (i.e. 'perfect health'),<sup>(3)</sup> which seemed unlikely to accurately reflect their true health state, given its assessment of areas such as mobility and pain. However, proxy-assessors can have different perceptions of another's capabilities and health state,<sup>(2, 4)</sup> perhaps particularly for more unobservable or subjective dimensions (e.g. pain or anxiety/depression versus functional ability), and exhibit additional sources of bias which could differ between individuals. For example, a relative's response might reflect additional emotional burden when compared to formal/employed carers. Ethical issues arise if the views of individuals about their own HRQOL differ from the views of their proxies, especially if these affect economic evaluation outcomes and hence resource allocation decisions affecting them.

Dementia-specific PBMs (e.g. DEMQOL-U or AD-5D) are designed to capture physical function and cognitive changes that are common in dementia, but poorly reflected in generic measures <sup>(5-8)</sup> and have been valued by the general (UK) population.<sup>(6)</sup> Dementia researchers thus face a relatively unique challenge in choosing an appropriate measure (generic vs. condition-specific) as well as respondent (self vs. informal carer vs. formal carer).

This paper's main purpose is to examine and compare the measurement properties of proxy- versus self- and generic- versus dementia-specific PBMs. Analyses of construct validity, criterion validity and responsiveness

will aid decisions about how best to capture utility (and change in utility) and thus support the interpretation and design of health economic analyses involving people with dementia.

This study used a large, three wave, individual-level dataset of people with dementia living in care homes collected during a multi-site trial (NAME REDACTED).(9) The data includes a rich combination of resident-completed, proxy-reported, generic- and dementia-specific PBMs as well as other non-PBM cognitive and HRQOL measures. Whilst an Australian (4, 10) study has compared the included PBMs, this UK based study used larger sample sizes, more cognitive health and HRQOL measures, and involved modelling temporal within-individual changes, using panel data regression techniques.

## 2. Methods

### 2.1 Dataset

Residents (n=1,004) were recruited from fifty UK care homes in an open cohort cluster RCT (TRIAL NAME REDACTED) which evaluated (INTERVENTION NAME REDACTED), an intervention that aims to reduce agitation and improve HRQOL for people with dementia.(9) Residents (supported by researchers), a formal-carer proxy-assessor (member of care home staff) and, where possible, a further informal-carer proxy-assessor (a friend or relative who visited at least fortnightly) were invited to complete PBMs and other questionnaires at one (n=425), two (n=173) or three (n=406) time points (when they entered the trial and at up to two follow-ups) between May 2014 and May 2017. The mean time between time points was 252 days (s.d. 55.9). Not all residents were assessed at all three time points due mainly to death (n=281), recruitment after trial baseline (n=261) or moving away (n=47) (See Figure A1). Formal-carer proxies were replaced whenever they became unavailable (e.g. due to staff turnover, sickness or other absence). Not all residents had an informal-carer proxy due to lack of regular visitors who wished to participate and their recruitment ceased prior to the third time point.

## 2.2 Description of variables

Table 1 summaries the resident- and proxy-completed questionnaires used in this study. Residents' health status was self-reported using the five-dimension EQ-5D-5L) and proxy-reported using the EQ-5D-5L proxy (( version 1)(11) and the 31-item DEMQOL-Proxy. UK scoring algorithms were used to convert these to the EQ-5D-5L and DEMQOL-Proxy-U (using four items from the DEMQOL-Proxy, namely appearance, memory, negative and positive emotion) utility scores.(6)

Residents' degree of cognitive and functional impairment was assessed by formal carer-proxies using:

Functional Assessment Staging (FAST), which scores functional symptoms of dementia from 1 ('normal adult' without difficulties) to 7 ('severe dementia,' e.g. daily speech limited to  $\leq 5$  words).(12) In this study, scores  $\leq 4$  (i.e. 'mild dementia,' e.g. difficulties planning dinner for guests) were combined into a single category since recruitment to the trial required a score  $\geq 4$ .

Clinical Dementia Rating Scale (CDR), which assesses cognitive impairment in six categories (e.g. related to memory or orientation) and generates a single severity score ranging from 0 ('normal adult') to 3 ('severe dementia).(13, 14)

Cohen Mansfield Agitation Inventory (CMAI), which measures the frequency of 29 behaviours typically associated with agitation on a seven-point scale (from 'never' to 'several times an hour') and generates a single score ranging from 29 to 203 (where changes  $\geq 8$  can be considered clinically meaningful).(15-18)

Residents' HRQOL was self-reported using the 15-item nursing home version of the Quality of Life in Alzheimer's Disease scale (QOL-AD-NH), which scores items including energy, mood, living situation and memory on four-point scales, and by both proxies using the 11-item Quality of Life in Late Stage Dementia measure (QUALID), which scores items including smiling, sadness and enjoyment of interaction with others on six-point scales.(19) In both cases, the sum of item-level scores generates a single measure on a 45-point scale.

## 2.3 Statistical analysis

### 2.3.1 Descriptive analysis

Summary statistics (including the extent of missing data) were reported for variables of interest. Selected frequency distribution were examined using histograms.

### 2.3.2 Construct validity

Construct validity was examined by comparing each PBM-respondent utility score. First, pairs of utility scores generated by different PBMs completed by the same proxy respondent at the same time point were compared (e.g. informal carers' EQ-5D-5L-proxy and DEMQOL-Proxy-U responses) using Spearman's rank order correlation (where correlation  $<0.3$  was considered weak,  $0.3$  to  $<0.5$  moderate,  $0.5$  to  $<0.6$  strong, and  $\geq 0.6$  very strong,(4)) and Bland–Altman plots (differences between two utility values generated by two different PBMs plotted against the mean of the utility values). Second, utility scores generated from the same PBM completed by different respondents at the same time point were compared (e.g. informal-carer EQ-5D-5L-proxy versus formal-carer EQ-5D-5L-proxy or resident-completed EQ-5D-5L) using Spearman's rank order correlation and Bland–Altman plots (first, differences between two utility values recorded at the same time point were plotted against the mean, and second, differences between the change in utility values between two time points were plotted against the mean at the first time point).

### 2.3.3 Criterion validity

Criterion validity was examined by comparing utility scores and non-utility measures. Associations between each PBM and the non-PBM cognitive (FAST, CDR and CMAI) and HRQOL measures (QUALID and QOL-AD-NH) (described above in 2.2) were examined at the same time point using Spearman's rank order correlation and panel data regression models with random effects (model 1). In the absence of evidence that other regression models would perform better in a dementia population, all regression analyses in this study used linear model specifications. This enabled model output comparisons across multiple outcome measures.



Model 1:  $PBMscore_{i,t} = \beta_1 + \beta_2 health_{i,t} + \epsilon$

*PBMscore<sub>i,t</sub> is the utility score for each individual (i) at each time point (t) for a particular PBM-respondent combination (n=5)*

*health<sub>i,t</sub> is a particular cognitive or HRQOL measure (n=6). This was a single continuous variable, except for the FAST and CDR scores where a vector of categorical variables was created (with least impaired states in the reference category).*

The model was run separately for each PBM-respondent combination (n=5) and cognitive or HRQOL variable (n=6). Missing data for particular utility scores meant sample sizes varied in each analysis. A sensitivity analysis used a restricted sample of observations where all five utility scores were completed.

The coefficient of interest ( $\beta_2$ ) shows the association between an additional unit of cognitive health or HRQOL (e.g. a one unit change in CMAI), or being in a particular health category compared to the reference category (e.g. FAST score 6 compared to FAST score  $\leq 4$ ), and a particular utility score. 95% confidence intervals were calculated and adjusted for multiple comparisons using Bonferroni correction (n=6 per outcome measure). Statistically significant  $\beta_2$  values indicated criterion validity.

#### 2.3.4 Responsiveness

The impact of changes in the cognitive or HRQOL measures (n=6) on each utility score (n=5) were evaluated using linear fixed effects panel data regression models (as model 1 above). The fixed effects approach means that only within-individual changes are included in the analysis. Hence statistically significant  $\beta_2$  values were deemed indicative of responsiveness.

### 3. Results

### 3.1 Descriptive statistics

The sample included 1,989 observations from 1,004 residents and an average of 1.98 time points per resident (SD: 0.91). Of these, 406 residents (40.4%) participated in the study at all three time points, whilst the remainder were in the study for only one (42.3%) or two (17.2%) time points.

Residents ranged from 58.0 to 102.6 years of age on entering the dataset (mean=85.5 years) and were more likely female (73.2%) than male. 3.3% had ethnic backgrounds other than White British (see Table 2 for descriptive statistics, with selected frequency distributions shown in Figure A2).

#### 3.1.1 PBMs

Sample sizes were largest for the formal-carer EQ-5D-5L-proxy and DEMQOL-Proxy-U, with at least one observation for 1,003 individuals (100%) (N=1,980 for EQ-5D-5L-proxy and N=1,983 for DEMQOL-Proxy-U). Missing formal carer-completed PBM data was rare (15 instances) and due to partial questionnaire completion (1-2 missing items, 13 instances) or non-completion (2 instances).

At least one resident-reported EQ-5D-5L utility score was available for 558 individuals (56%) in the dataset (N=897, 45%) and, of these, an average of 1.60 observations per individual. There were 67 instances of partial completion (1-4 missing items) (3.4%) and 1,025 instances (52%) of non-completion (with reasons recorded in 1,006 instances, typically related to the resident being unable or reluctant to communicate). The likelihood of non-completion increased with dementia severity, rising for example from 17% of observations in the 'Normal or Very Mild Dementia' CDR category to 83% in the 'Severe dementia' category.

Sample sizes were smallest for the informal-carer PBMs (n=184 (18.3% of individuals), EQ-5D-5L-Proxy: N=349, DEMQOL-Proxy-U: N=342 (17.5% of observations)) due mainly to fewer eligible informal-carer proxies being recruited, but also due to 1-3 missing items on the PBM questionnaire (33 instances).

The mean value of DEMQOL-Proxy-U scores were higher than the EQ-5D-5L-proxy scores (ranging from 0.486 for informal-carer EQ-5D-5L-proxy to 0.751 for formal-carer DEMQOL-Proxy-U), with a lower standard

deviation. The resident-completed EQ-5D-5L had the highest mean utility score (0.861). Across all PBM-responder combinations, within-individual deviation was smaller than between-individuals.

### 3.2 Construct validity

Figures A3-A5 show the Spearman's rank order correlation coefficients and Bland-Altman plots. Figures A6-A9 show comparisons between over-time changes in the resident-completed and two proxy-completed EQ-5D-5L utility scores.

#### 3.2.1 Different PBMs completed by the same proxy-responder

Correlation between informal carer-completed utility scores was moderate ( $\rho=0.386$ ) and between formal-carer completed utility scores was weak (0.191) (Figure A3, (i.) and (ii.)). For both proxies, the DEMQOL-Proxy-U scores were higher on average than EQ-5D-5L-proxy scores, although discrepancies between scores were lower at higher utility levels.. Mean differences between EQ-5D-5L-proxy and DEMQOL-Proxy-U utility scores were greater amongst informal-carer (0.214) than formal-carer (0.092) proxies, with smaller discrepancies at higher utility levels.

#### 3.2.2 Different respondents completing the same PBM

Correlation between informal- and formal-carer proxy-completed utility scores was strong for EQ-5D-5L ( $\rho=0.528$ ) and weak for DEMQOL-Proxy-U (0.088) (Figure A3, (iii.) and (iv.)). For both PBMs, formal-carer proxy-completed scores were higher on average than the informal-carer proxy-completed scores, although for EQ-5D-5L they were lower at higher utility levels and for DEMQOL-Proxy-U the discrepancies between scores were very small, especially at the highest and lowest utility levels).

Correlation between formal-carer EQ-5D-5L-proxy and resident-reported EQ-5D-5L utility scores was weaker ( $\rho=0.264$ ) than that between informal-carer EQ-5D-5L-proxy and resident-completed EQ-5D-5L utility scores (0.354) (Figure A4 and A5). In both cases, correlation ( $\rho$ ) between the carer-reported EQ-5D-5L-proxy and resident-reported EQ-5D-5L utility scores decreased, and mean differences increased, with dementia severity (CDR score). On average, EQ-5D-5L-proxy scores were lower than the resident-completed EQ-5D-5L scores (e.g. mean difference of -0.118 for formal-carers and -0.275 for informal-carers), with the greatest discrepancies between utility scores at lower utility levels. On average, over-time changes in utility were smaller for EQ-5D-5L-proxy respondent scores when compared to resident-reported EQ-5D-5L scores, although there was considerable heterogeneity between individuals (Figures A6-A9).

### 3.3 Criterion validity

Table A1 shows the Spearman's rank order correlation results. . Figure 1 shows regression results and Figure A10 shows the sensitivity analysis where data was complete for all five utility scores.

#### 3.3.1 Cognitive measures

##### 3.3.1.1 Relationship with resident-completed EQ-5D-5L scores

No relationship was identified in either the Spearman's rank or (Bonferroni adjusted) regression analysis between resident-completed EQ-5D-5L and the CDR or FAST measures. Although prior to Bonferroni adjustment a statistically significant relationship was observed between the resident-completed EQ-5D-5L and the CMAI measure, the effect size was very small (an eight-unit increase in CMAI was associated with lower utility of 0.01).

##### 3.3.1.2 Relationship with proxy-completed EQ-5D-5L scores

The relationship between both EQ-5D-5L-proxy utility scores and the FAST and CDR measures were shown by the Spearman's rank coefficients to be moderate (ranging from  $\rho = -0.346$  to  $-0.492$ ) and, in the regression results, to be statistically significant and negative (with one exception). All regression coefficients increased in magnitude with dementia severity (e.g. for formal-carers, the coefficient for FAST=7 was  $-0.323$  compared with  $-0.177$  for FAST=6, both compared to FAST $\leq 4$ ). The size of the coefficients were generally comparable for both proxies (e.g. coefficient for FAST 6 compared to FAST 4 was  $-0.18$  in formal carer analysis and  $-0.16$  in informal carer analysis).

The relationship between the formal-carer EQ-5D-5L-proxy utility scores and the CMAI measure was statistically significant in the regression results with larger coefficients than in the analyses of other utility measures, including the analyses of informal-carer EQ-5D-5L-proxy utility scores where coefficients were not statistically significant.

#### 3.3.1.3. Relationship with DEMQOL-Proxy-U scores

The relationship between both proxy-completed DEMQOL-Proxy-U utility scores and the FAST and CDR measures were very weak in the Spearman's rank analysis. The only statistically significant effects observed in the regression analysis were the relationship between formal-proxy DEMQOL-Proxy-U and the CDR score, however the effect sizes were notably smaller than those observed in the analyses of EQ-5D-5L-proxy scores completed by the same proxy.

As with the EQ-5D-5L-proxy analyses, the relationship between the CMAI score and the formal-carer DEMQOL-Proxy-U score was of a higher magnitude than the informal-carer DEMQOL-Proxy-U score, which was not statistically significant.

#### 3.3.2 Other HRQOL measures

The relationship between resident-reported QOL-AD-NH and EQ-5D-5L was weak in the Spearman's rank analysis ( $\rho=0.28$ ) and in the regression analysis it was statistically significant. This contrasted with the finding of no relationship observed in the Spearman's rank analysis, and no statistically significant relationship observed in the regression analysis, between (resident-reported) QOL-AD-NH and the other four utility measures.

Conversely, in the Spearman's rank analysis, the formal-carer completed QUALID had a weak to moderate relationship with all utility scores, except for the resident-reported EQ-5D-5L. The informal-carer completed QUALID also had a weak to moderate relationship with both the informal-carer EQ-5D-5L-proxy and DEMQOL-Proxy-U.

In the regression analysis, the QUALID scores had a statistically significant relationship with all utility scores. The magnitude of the relationship between QUALID and a utility score was always greater when the QUALID and utility score were completed by the same proxy. For example, a ten-unit change in formal-carer QUALID was associated with a 0.112 change in the formal-carer EQ-5D-5L-proxy and a 0.083 change in the informal-carer EQ-5D-5L-proxy.

### 3.4 Responsiveness

Figures 2 and A11 show the coefficients from the regression models, outlined below.

#### 3.4.1 Cognitive measures

##### 3.4.1.1. Relationship with resident-completed EQ-5D-5L

No statistically significant relationships were observed between changes in the FAST, CDR or CMAI measures and changes in the resident-completed EQ-5D-5L.

#### 3.4.1.2. Relationship with formal proxy-completed PBMs

A statistically significant (negative) relationship was observed between changes in both the CMAI and CDR measures, and changes in the formal-carer PBMs (EQ-5D-5L-proxy and DEMQOL-Proxy-U). The formal-carer EQ-5D-5L-proxy also captured a change to FAST score 6 and FAST score 7 (indicating more cognitive and functional impairment), however the formal-carer DEMQOL-Proxy-U was not responsive to any changes in FAST scores.

#### 3.4.1.3. Relationship with informal proxy-completed PBMs

Statistically significant (negative) relationships were observed between all possible changes in the FAST score and the informal-carer EQ-5D-5L-proxy and DEMQOL-Proxy-U (with one exception). In the analyses of FAST scores, the magnitude of the relationships with informal-carer PBMs was always greater than the formal-carer PBMs, whereas for the CDR and CMAI analyses, the coefficients were always larger for the formal-carer PBM.

#### 3.4.2 Other HRQOL measures

A statistically significant relationship was always observed between the HRQOL measures and PBMs which had been completed by the same respondent (e.g. the relationship between changes in resident-reported QOL-AD-NH and resident-reported EQ-5D-5L was statistically significant).

The relationship between changes in the (proxy-completed) QUALID score and changes in the EQ-5D-5L-proxy were also always statistically significant, regardless of which proxy had completed the assessment, but this was not the case in the analyses of the relationship between QUALID and DEMQOL-Proxy-U.

## 4. Discussion

This study identified disparities in the measurement properties of two different PBMs completed by up to three different respondents (residents, formal-carer and informal-carer proxy-assessors) about one individual.

### 4.1 How do proxy-completed PBMs compare to resident-completed EQ-5D for people with dementia?

Consistent with existing studies, we found the likelihood of eliciting a self-completed EQ-5D response decreased substantially with dementia severity, with non-response rising to four-fifths of residents with 'severe dementia' (FAST=7). In contrast, EQ-5D-5L-proxy and DEMQOL-Proxy-U data collection via a formal carer proxy was feasible for almost all residents, demonstrating the role of such proxies in eliminating a substantial risk of sample bias if only resident-reported measures were used. However, informal carer proxies could only be recruited for a minority of residents (<20%) indicating that for a care home-based sample of people with dementia, staff member proxies would likely provide the largest number of responses.

Our finding that resident-completed EQ-5D-5L scores were of a greater magnitude than proxy-completed EQ-5D-5L scores was also consistent with existing studies of dementia (21-27) and some other conditions.(21) Researchers should consider any influence this might have on the results of a trial using proxy-completed data.

In tests of criterion validity and responsiveness, which NICE explicitly recommends for determining the appropriateness of EQ-5D for a particular population,(1) the self-completed EQ-5D-5L was poor at reflecting clinically important differences and changes in the FAST, CDR and CMAI scores, but rather better at capturing (the resident's own view of) quality of life (QOL-AD-NH). This is consistent with some comparable studies involving people with dementia that identified correlation between resident-reported EQ-5D and QOL-AD-NH,(28-30) but not between resident-reported EQ-5D and some cognitive measures (e.g. PAS-COG or NPI-Q).(4, 22, 28, 30).

Our finding that the EQ-5D-5L-proxy and DEMQOL-Proxy-U did capture changes in the FAST, CDR and CMAI may justify their use in research, especially if the intervention under investigation is expected to affect these more objective, cognitive aspects of health. Nevertheless, we also found the formal- and informal-carer



completed EQ-5D-5L-proxy and DEMQOL-Proxy-U were unable to capture the resident's perception of their own quality of life (as measured by QOL-AD-NH), despite adequately capturing the proxies own view of a resident's quality of life (QUALID). This might reflect discrepancies between proxy-views and resident-views of the impact that a tangible decline in cognition, functional abilities or clinical aspects of health has on quality of life. Perhaps the resident did not notice that decline, or they felt it did not unduly interfere with their quality of life. For instance, despite declining physical mobility, residents might still enjoy full access to the things they value because of support from staff and/or aids (e.g. wheelchairs or hoists). Their reference point, when considering what is meant by 'usual activities' for example, may also differ from their proxy who might more likely think about younger, healthier people in the general population than fellow residents in the care home. Hence, without a fuller understanding of the degree to which residents cognitive impairments had prevented them from being able to appreciate their current situation reliably, researchers (and NICE) face a considerable challenge in terms of deciding when it is desirable to give less weight to a resident's own views about what constitutes HRQOL. After all, this could lead to interventions deemed of value by proxies being prioritised over other interventions that would have greater impact on those aspects of HRQOL residents themselves say they value the most.

#### 4.2 If researchers opt for proxy-completed PBMs, does it matter which they use?

We found that EQ-5D-5L-proxy and DEMQOL-Proxy-U were generally weak or poor substitutes for measuring utility. Regardless of which proxy completed it, EQ5D-5L-proxy was typically more responsive than DEMQOL-Proxy-U to changes in CDR, FAST and CMAI. This probably reflects the relative content of the measures: EQ-5D examines the frequency and severity of symptoms and functional impairment directly, whereas DEQMOL-Proxy-U has a broader focus on the emotional impact of dementia across wider QOL domains. Nevertheless, since DEMQOL-Proxy-U was no better than the EQ-5D-5L-proxy at reflecting aspects of resident-reported HRQOL included in the QOL-AD-NH, this study provides no substantive evidence to justify using DEMQOL-Proxy-U over the default (and perhaps more straightforward(33)) EQ-5D-Proxy.

We also identified moderate correlation between the informal carer-completed EQ-5D-5L-proxy and DEMQOL-Proxy-U, with DEMQOL-Proxy-U providing higher utility scores especially at lower levels. These results closely resemble the two comparable Australian studies (e.g. we reported  $\rho=0.386$ , versus 0.389 in one

study, and a mean difference of 0.214, versus 0.202)(4, 10) and might have arisen because the EQ-5D-5L has a much lower bound than the DEMQOL (-0.285 vs 0.363, respectively).

#### 4.3 Comparison with existing studies

Three existing (repeated-) cross-sectional studies compared various combinations of resident-completed and proxy-completed versions of EQ-5D and DEMQOL in Britain or Australia. In contrast to our study, these included people without dementia(4) or were restricted to unrepresentative subgroups (e.g. people recently hospitalised(10) or with significant depressive symptoms). None compared formal- with informal-carer proxies, nor DEMQOL-Proxy-U with other quality of life measures (QOL-AD-NH and QUALID), and two did not compare self- and proxy- responses for the same residents.

Our study uniquely collected two dementia-specific scales of cognitive and functional impairment (FAST and CDR), rather than more general cognition or mental health measures (e.g. MMSE or PAS-COG), and included residents with the full range of dementia severity. Larger sample sizes, repeated measures and/or longer follow-up periods also enabled analyses of responsiveness that likely support more robust causal inference than cross-sectional comparisons.

#### 4.4 Limitations

Some caution is necessary when comparing person-specific utility scores measured at the same time point because each assessor completed the PBM questionnaires at different times within a two week window, and because by design the EQ-5D-5L (which measures ‘health today’) and DEMQOL (‘the past week’) assess HRQOL over different time periods (Table 1).

Although this study included multiple utility measures for each resident, it did not include (resident-reported) DEMQOL-U (due to concerns about research burden) nor broader capability measures (e.g. ICECAP-O(34)).

#### 4.5 Implications and future research

This study can inform decisions about how to measure utility in studies of people with dementia, depending for example on the functional or cognitive aspects of HRQOL researchers wish to capture, as well as the severity of dementia in trial participants. A future study could collect the quantitative PBM data alongside immediate qualitative interviewing to better understand reasons for observed discrepancies between resident- and proxy-rated utility. Given the difficulties of collecting utility data from residents (when they were unable) and proxies (because of recruitment challenges) in this study, as well as discrepancies between utility measures, trialists would be well advised to try to collect PBM data from both individuals and their proxies. Future work could establish how to improve the participation of proxies in trials and develop more specific guidance on when particular utility measures should be used, based for instance on dementia severity. This requires more information on the degree to which discrepancies between proxy- and resident-reported outcomes arise because residents are no longer able to judge their own situation reliably, for instance, and on which particular improvements in functional or cognitive aspects of HRQOL are most valued by payers. A potential focus of future research could be to combine different utility scores to provide a more complete utility measure that utilises the insights provided by residents as well as their proxies, by ‘adjusting’ the scores of one using the score provided by another.

The observed relationships between the utility scores and (non-PBM) cognitive measures could also be used to inform health state utility parameters associated with several different clinical measures in economic models, by indicating the utility decrement associated with different (worsening) dementia severity, and to inform imputation techniques in trials where some utility measures are missing.

#### 5. Conclusion:

There are multiple strategies to capture health state utility in people with dementia. Choice of strategy has a non-trivial impact on the utility values obtained and these will potentially impact on economic evaluation results.

The results presented here can inform choice of utility capture strategy. Future research should explore how to combine utility data from multiple different sources to prevent data loss and to retain the patient's perspective.



## References

1. National Institute for Clinical Excellence. Guide to the methods of technology appraisal. 2013.
2. Diaz-Redondo A, Rodriguez-Blazquez C, Ayala A, et al. EQ-5D rated by proxy in institutionalized older adults with dementia: Psychometric pros and cons. *Geriatrics & Gerontology International*. 2014; 14: 346-53.
3. Coucill W, Bryan S, Bentham P, et al. EQ-5D in patients with dementia: an investigation of inter-rater agreement. *Medical care*. 2001; 39: 760-71.
4. Easton T, Milte R, Crotty M, et al. An empirical comparison of the measurement properties of the EQ-5D-5L, DEMQOL-U and DEMQOL-Proxy-U for older people in residential care. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2018.
5. Gomes M, Pennington M, Wittenberg R, et al. Cost-effectiveness of Memory Assessment Services for the diagnosis and early support of patients with dementia in England. *Journal of Health Services Research & Policy*. 2017; 22: 226-35.
6. Rowen D, Mulhern B, Banerjee S, et al. Estimating preference-based single index measures for dementia using DEMQOL and DEMQOL-Proxy. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2012; 15: 346-56.
7. Comans TA, Nguyen K-H, Mulhern B, et al. Developing a dementia-specific preference--based quality of life measure (AD-5D) in Australia: a valuation study protocol. *BMJ Open*. 2018; 8: e018996.
8. Connolly S. Economics of dementia: A review of methods. *Dementia*.
9. REFERENCE REDACTED.
10. Ratcliffe J, Flint T, Easton T, et al. An Empirical Comparison of the EQ-5D-5L, DEMQOL-U and DEMQOL-Proxy-U in a Post-Hospitalisation Population of Frail Older People Living in Residential Aged Care. *Applied health economics and health policy*. 2017; 15: 399-412.
11. Dolan P. Modeling valuations for EuroQol health states. *Medical care*. 1997; 35: 1095-108.
12. Reisberg B. Functional assessment staging (FAST). *Psychopharmacology Bulletin*. 1988; 24: 653-59.
13. Morris JC. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*. 1993; 43: 2412-14.
14. Hughes CP, Berg L, Danziger WL, et al. A new clinical scale for the staging of dementia. *The British Journal of Psychiatry*. 1982; 140: 566-72.
15. Zuidema SU, Buursema AL, Gerritsen MG, et al. Assessing neuropsychiatric symptoms in nursing home patients with dementia: reliability and Reliable Change Index of the Neuropsychiatric Inventory and the Cohen-Mansfield Agitation Inventory. *International journal of geriatric psychiatry*. 2011; 26: 127-34.
16. Cohen-Mansfield J, Billig N. Agitated behaviours in the elderly. I. A conceptual review. *Journal of the American Geriatrics Society*. 1986; 34: 711-21.
17. Cohen-Mansfield J. Instruction manual for the Cohen-Mansfield agitation inventory (CMAI). Maryland: The Research Institute of the Hebrew Home of Greater Washington, 1991.
18. Cohen-Mansfield J, Marx MS, Rosenthal AS. A description of agitation in a nursing home. *The Journal of Gerontology*. 1989; 44: M77-M84.
19. Weiner MF, Martin-Cook K, Svetlik DA, et al. The Quality of Life in Late-Stage Dementia (QUALID) Scale. *Journal of the American Medical Directors Association*. 2000; 1: 114-16.
20. REFERENCE REDACTED.
21. Rand S, Caiels J. Using proxies to assess quality of life: a review of the issues and challenges. 2015.
22. Kunz S. Psychometric properties of the EQ-5D in a study of people with mild to moderate dementia. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2010; 19: 425-34.
23. Mulhern B, Rowen D, Brazier J, et al. Development of DEMQOL-U and DEMQOL-PROXY-U: generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. *Health technology assessment (Winchester, England)*. 2013; 17: 1-140.
24. Naglie G, Tomlinson G, Tansey C, et al. Utility-based Quality of Life Measures in Alzheimer's Disease. *Quality of Life Research*. 2006; 15: 631-43.
25. Devine A, Taylor SJC, Spencer A, et al. The agreement between proxy and self-completed EQ-5D for care home residents was better for index scores than individual domains. *Journal of Clinical Epidemiology*. 2014; 67: 1035-43.
26. Moyle W, Murfield JE, Griffiths SG, et al. Assessing quality of life of older people with dementia: a comparison of quantitative self-report and proxy accounts. *Journal of Advanced Nursing*. 2012; 68: 2237-46.

27. Jonsson L, Andreasen N, Kilander L, et al. Patient- and proxy-reported utility in Alzheimer disease using the EuroQoL. *Alzheimer disease and associated disorders*. 2006; 20: 49-55.
28. Karlawish JH, Zbrozek A, Kinoshian B, et al. Preference-based quality of life in patients with Alzheimer's disease. *Alzheimer's and Dementia*. 2008; 4: 193-202.
29. Thorgrimsen L, Selwood A, Spector A, et al. Whose quality of life is it anyway? The validity and reliability of the Quality of Life - Alzheimer's Disease (QOL-AD) Scale. . *Alzheimer disease and associated disorders*. 2003; 17: 201-8.
30. Hounscome N, Orrell M, Edwards RT. EQ-5D as a Quality of Life Measure in People with Dementia and Their Carers: Evidence and Key Issues. *Value in Health*. 2011; 14: 390-99.
31. National Institute for Clinical Excellence. Guide to the methods of technology appraisal. 2008.
32. Brazier J, Longworth L. NICE Decision Support Unit Technical Support Documents. NICE DSU Technical Support Document 8: An Introduction to the Measurement and Valuation of Health for NICE Submissions. London: National Institute for Health and Care Excellence (NICE), 2011.
33. McCartney K, Robinson K, Horne JC, et al. DEMQOL: Fit for purpose in care homes? Age and Ageing. 2018; 47: iii9-iii12.
34. van Leeuwen KM, Bosmans JE, Jansen APD, et al. Comparing Measurement Properties of the EQ-5D-3L, ICECAP-O, and ASCOT in Frail Older Adults. *Value in Health*. 2015; 18: 35-43.

Table 1: Description of questionnaires used in the study

Description of questionnaire							Description of variable used in this analysis	
	What it measures	Resident	Formal carer proxy	Informal carer proxy	Perspective adopted by proxies	Timescale	Conversion	Possible values
Preference based measures (PBMs) (increasing in quality of life)								
EQ-5D-5L	Five dimensions: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression	✓	✓ (*)	✓ (*)	Their own opinion of the resident's health status	Resident's immediate situation - i.e. their 'health today'	Converted to an index utility score using a general population valuation of the health states	-0.285 to 1
DEMQOL-PROXY	31 items including factors related to appearance, memory, positive and negative emotions	✗	✓	✓	Their view of what they think residents would provide themselves if they were willing and able	How the resident felt during the past week	Relevant components of the DEMQOL-Proxy converted to an index utility score (DEMQOL-Proxy-U) using a general population valuation of the health states	0.363 to 0.937
Cognitive measures (increasing in severity of dementia or agitation)								
FAST	Functional severity of dementia on a scale from 1 (a 'normal adult' with no difficulties) to 7 ('severe dementia')	✗	✓	✗	Not specified	Not specified	FAST scores ≤4 were combined into a single category (due to small sample sizes where FAST score <4)	Four categories: ≤4, 5, 6, 7
CDR	Cognitive impairment on six cognitive categories	✗	✓	✗	Not specified	Not specified	Converted to an overall severity rating ranging from 0 to 3. Ratings of 0, 0.5 and 1 were combined in this analysis (due to small sample sizes where CDR<1)	Three categories: ≤1, 2, 3
CMAI	Frequency of 29 agitated or aggressive behaviours on a seven-point scale ranging from 1 ('never') to 7 ('several times an hour')	✗	✓	✗	Not specified	Based on observations during the previous fortnight	Frequency scores are summed	29 to 203
Other HRQOL measures (increasing in quality of life)								
QOL-AD-NH	15 aspects of quality of life including perception of energy, mood, living situation and memory on a four-point scale (poor, fair, good, excellent)	✓	✗	✗	N/A	N/A	Individual scores are summed	45-point scale
QUALID	Presence and frequency of eleven quality of life indicators including smiling, appearing sad, crying and enjoying eating	✗	✓	✓	Not specified	Based on observations during the previous week	Individual scores are summed	45-point scale (**)

### Notes

\*indicates EQ-5D-5L proxy version 1

\*\*To aid interpretation and for consistency with the QOL-AD-NH, but contrary to general usage, the QUALID scores were reversed in this study so that higher scores represented improved quality of life.

Except for FAST and CDR, all variables were treated as continuous variables.



Table 2: Descriptive statistics for PBMs and other cognitive and HRQOL measures

.	Number of observations (N) and % of all	Number of individuals (n) and % of all	Number of observations per individual	Mean value	Range	Variation		
						Overall (i.)	Between individuals (ii.)	Within individuals (iii.)
PBMs (INCREASING IN QUALITY OF LIFE)								
EQ-5D						(a)		
Resident	897 (45.1%)	558 (55.6%)	1.607	0.861	-0.092 to 1	0.175	0.163	0.087
Formal-carer	1980 (99.5%)	1003 (99.9%)	1.974	0.658	-0.218 to 1	0.241	0.221	0.118
Informal-carer	349 (17.5%)	184 (18.3%)	1.897	0.486	-0.281 to 1	0.249	0.224	0.105
DEMQOL-Proxy-U						(a)		
Formal-carer	1983 (99.7%)	1003 (99.9%)	1.977	0.751	0.363 to 0.937	0.117	0.100	0.071
Informal-carer	342 (17.5%)	184 (18.3%)	1.858	0.701	0.404 to 0.937	0.127	0.112	0.062
COGNITIVE MEASURES (INCREASING IN SEVERITY OF DEMENTIA OR AGITATION)								
CMAI						(a)		
Formal-carer	1964 (98.7%)	985 (98.1%)	1.994	45.114	29 to 129	16.570	15.518	7.693
FAST (categorical variable)						(b)		
Formal-carer	1954 (98.2%)	994 (99.0%)	1.966	n/a	1 to 4	192 (9.83%)	162 (16.3%)	64.3%
					5	156 (7.98%)	146 (14.7%)	56.05%
					6	1165 (59.62%)	728 (73.2%)	82.07%
					7	441 (22.57%)	278 (28.0%)	75.72%
CDR (categorical variable)						(b)		
Formal-carer	1957 (98.4%)	987 (98.3%)	1.983	n/a	0 or 0.5	75 (3.8%)	70 (7.1%)	58.3%
					1	450 (23.0%)	341 (34.6%)	69.6%
					2	767 (39.2%)	539 (54.6%)	70.4%
					3	665 (34.0%)	433 (43.9%)	76.1%
OTHER HRQOL MEASURES (INCREASING IN QUALITY OF LIFE)								
QOL-AD-NH						(a)		
Resident	665 (33.4%)	448 (44.6%)	1.484	42.466	17 to 60	5.793	5.667	2.666
QUALID						(a)		
Formal-carer	1958 (98.4%)	987 (98.3%)	1.984	19.970	11 to 49	6.598	5.894	3.632
Informal-carer	360 (18.1%)	186 (18.5%)	1.935	22.162	11 to 44	7.499	6.955	3.211

- (a) For continuous variables, these three columns show standard deviation
- (b) For categorical variables (FAST and CDR), these three columns show:
  - i. number of observations in a given category, and as a percentage of all observations (total=100%)
  - ii. number of individuals that were ever in a given category (e.g. 70 individuals ever had CDR score=0 or 0.5), and as a percentage of all individuals (e.g. 7.1% of all individuals ever had CDR score=0 or 0.5)
  - iii. the fraction of time that individuals were in a given category, conditional on ever having been in that category (e.g. conditional on having ever had a CDR score=0 or 0.5, 58.3% of an individual's observations were in that category)

Figure 1: Associations between utility scores and non-utility cognitive and HRQOL measures (criterion validity)

See attached .EPS file

**Notes:**

Figure shows the  $\beta_2$  coefficients for model 1 run for each combination of utility (n=5) and non-utility cognitive measures (n=3).

The FAST and CDR scores are categorical variables. CDR 0, 0.5 and 1 are in the reference category. FAST  $\leq 4$  is in the reference category. The CMAI (range 29-203), QOLAD and QOL-AD-NH (45 point scales) scores are continuous variables.

The thicker confidence intervals indicate  $p < 0.05$ . The thinner confidence intervals are after adjustment for multiple hypothesis testing using Bonferroni correction.

Figure 2: Responsiveness of utility scores to changes in non-utility cognitive measures

See attached .EPS file

**Notes:**

Figure shows the  $\beta_2$  coefficients for model 1 run for each combination of utility (n=5) and non-utility cognitive measures (n=3).

The FAST and CDR scores are categorical variables. The  $\beta_2$  coefficient for FAST score 5 is compared here to the  $\beta_2$  coefficient for FAST score 4, coefficients for FAST score 6 are compared to FAST score 5, and FAST score 7 is compared to FAST score 6. Similarly, CDR score 2 is compared to the reference category (i.e. CDR 0, 0.5 and 1), and CDR score 3 is compared to CDR score 2. The CMAI (range 29-203) is a continuous variable.

The thicker confidence intervals indicate  $p < 0.05$ . The thinner confidence intervals are after adjustment for multiple hypothesis testing using Bonferroni correction.

The responsiveness of utility scores to changes in non-utility HRQOL measures are shown in Figure A11.

## Appendix

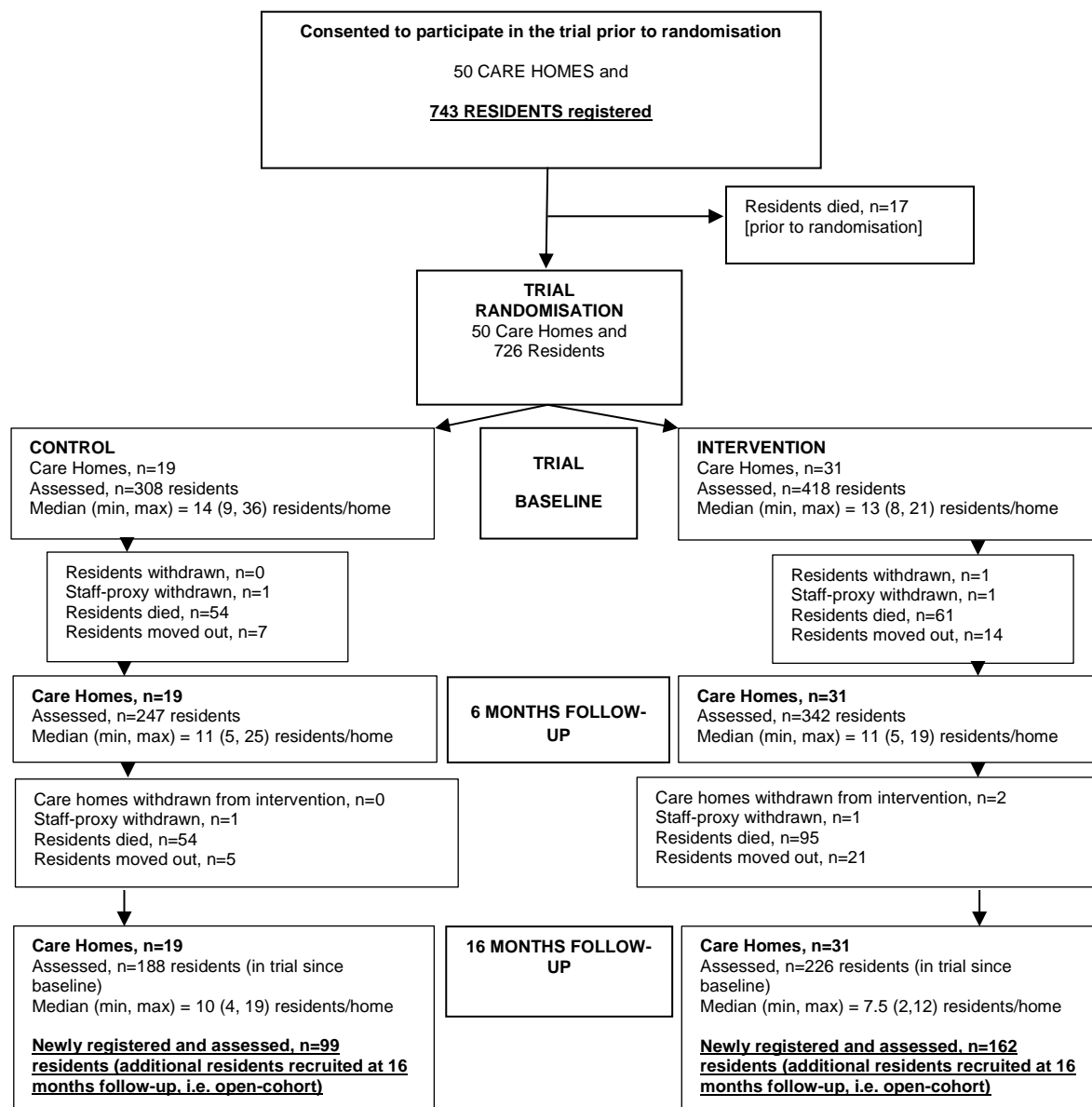
Table A1: Associations between utility and non-utility health measures (criterion validity)

Spearman's Rank Correlation Coefficients

	<b>Resident-reported EQ5D5L</b>	<b>Informal proxy-completed EQ5D5L</b>	<b>Formal proxy-completed EQ5D5L</b>	<b>DEMQOL-Proxy-U (informal-proxy)</b>	<b>DEMQOL-Proxy-U (formal-proxy)</b>
<b>FAST</b>	-0.0493 (N=888)	-0.3463*** (N=341)	-0.4917*** (N=1945)	-0.0044 (N=333)	-0.0510* (N=1948)
<b>CDR</b>	0.0250 (N=891)	-0.3371*** (N=344)	-0.4864*** (N=1948)	-0.0087 (N=337)	-0.1263*** (N=1951)
<b>CMAI</b>	-0.0663* (N=894)	0.0146 (N=345)	-0.0552* (N=1955)	-0.0008 (N=338)	-0.0693** (N=1958)
<b>QOL-AD-NH</b>	0.2817*** (N=649)	0.0635 (N=101)	0.1061** (N=662)	0.1253 (N=100)	0.0783* (N=664)
<b>QUALID (Formal-carer completed)</b>	-0.0969** (N=892)	-0.2705*** (N=343)	-0.3076*** (N=1949)	-0.1523** (N=336)	-0.3564*** (N=1952)
<b>QUALID (Informal-carer completed)</b>	-0.2227*** (N=159)	-0.4765*** (N=342)	-0.1954*** (N=359)	-0.4631*** (N=338)	-0.1522** (N=358)

Table shows Spearman's rank correlation coefficient (rho) and statistical significance test of Ho: PBM score and health/QOL measures are independent.

Figure A1: Residents included in the study of the validity and responsiveness of proxy- versus self- and generic- versus dementia-specific preference-based measures



A total of 1,004 residents were included in the analysis. Of these, n=17 had died prior to trial randomisation and so were not included in the trial, and n=261 were recruited as part of the open-cohort at the final trial follow-up point.

Figure A2: Frequency distribution of the utility scores for each PBM-respondent combination

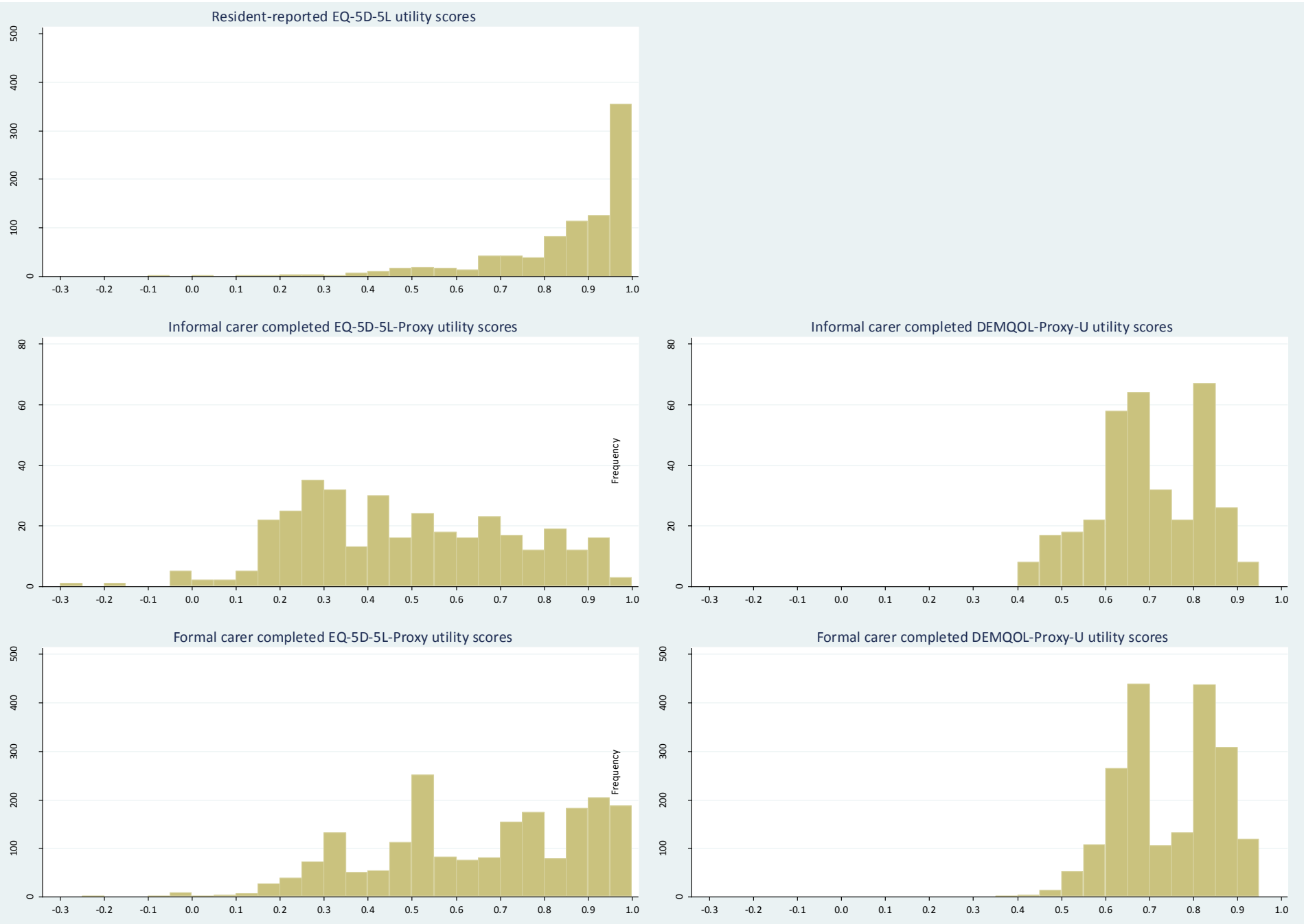
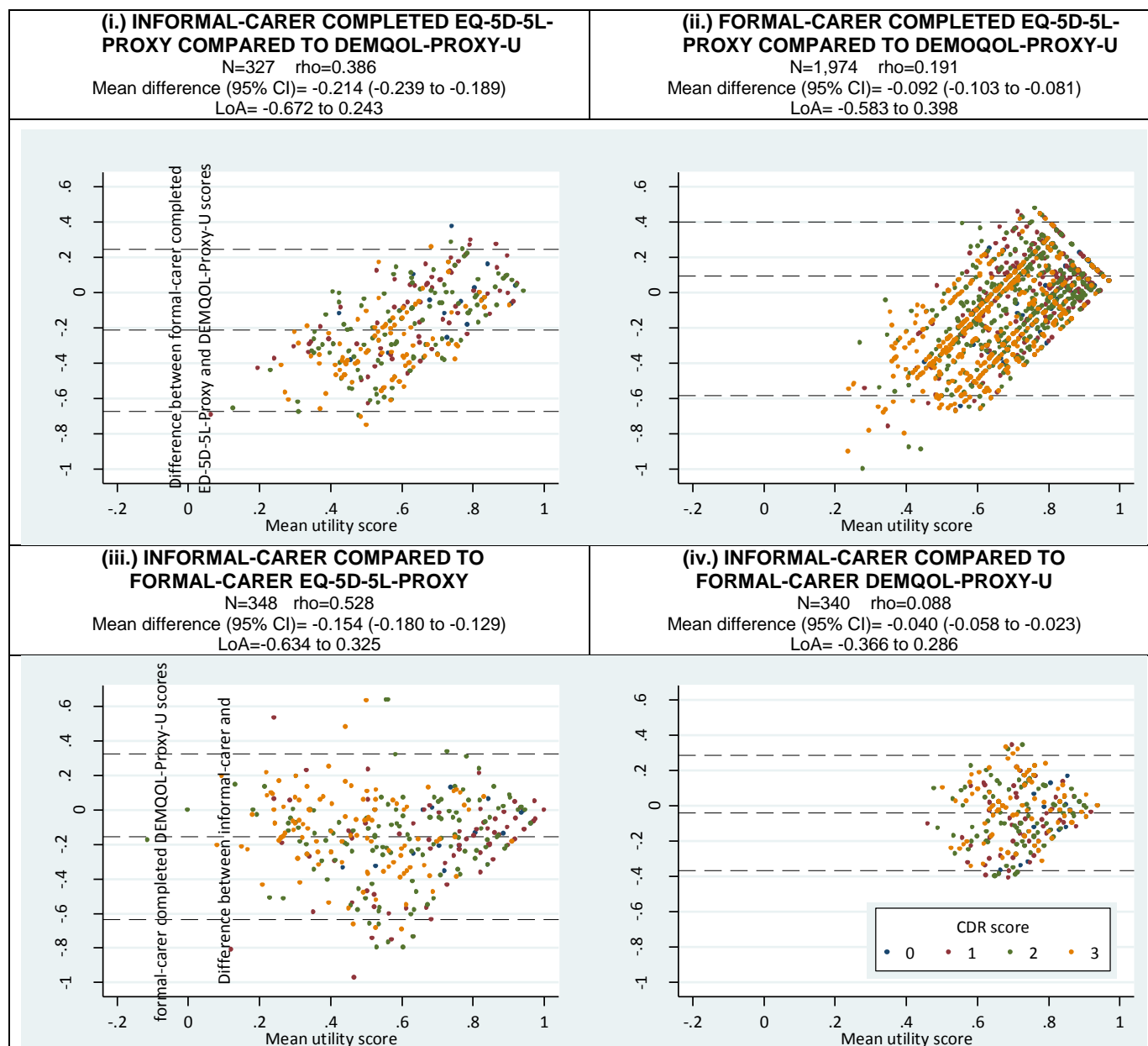


Figure A3: Differences between four pairs of proxy-completed utility scores



#### Notes:

Graphs show bland-Altman mean differences between four pairs of proxy-completed utility scores

N=number of person-time observations

rho: Spearman's rank correlation coefficient

Mean difference: Mean difference in utility scores. A mean difference value that differs significantly from 0 indicates the presence of a fixed bias.

LoA: Level of agreement (the mean difference plus and minus 1.96 times the standard deviation of the differences)

Horizontal lines are drawn at the mean difference and at LoA.



Figures A4 and A5:

Comparisons between resident-completed and two proxy-completed EQ-5D-5L utility scores, by dementia severity

**Notes:**

Graphs show Bland-Altman mean differences between resident-completed and two proxy-completed EQ-5D-5L utility scores

N: number of person-time observations

rho: Spearman's rank correlation coefficient

Mean difference: Mean difference in utility scores. A mean difference value that differs significantly from 0 indicates the presence of a fixed bias.

LoA: Level of agreement (the mean difference plus and minus 1.96 times the standard deviation of the differences)

Horizontal lines are drawn at the mean difference and at the LoA.

FIGURE A4:

INFORMAL-CARER COMPLETED EQ-5D-5L-PROXY COMPARED TO RESIDENT-COMPLETED EQ-5D-5L

	N	rho	Mean difference (95% CI)	LoA
Full sample	154	0.354	-0.275 (-0.318 to -0.232)	-0.811 to 0.261
Subsamples:				
CDR=0	14	0.551	-0.149 (-0.271 to -0.027)	-0.572 to 0.273
CDR=1	57	0.536	-0.217 (-0.282 to -0.152)	-0.705 to 0.272
CDR=2	73	0.291	-0.313 (-0.377 to -0.250)	-0.858 to 0.231
CDR=3	10	0.080	-0.507 (-0.702 to -0.312)	-1.052 to 0.039

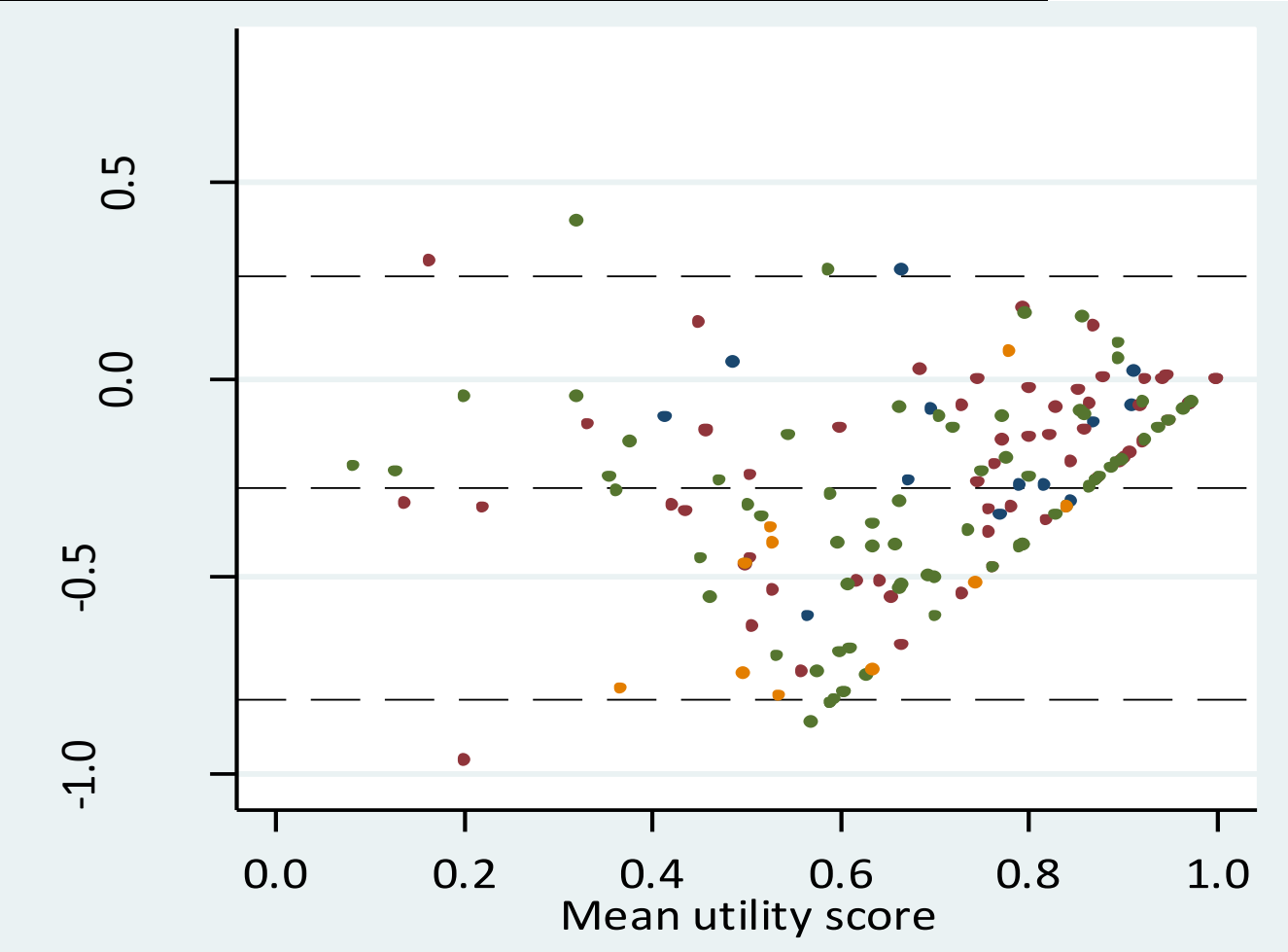
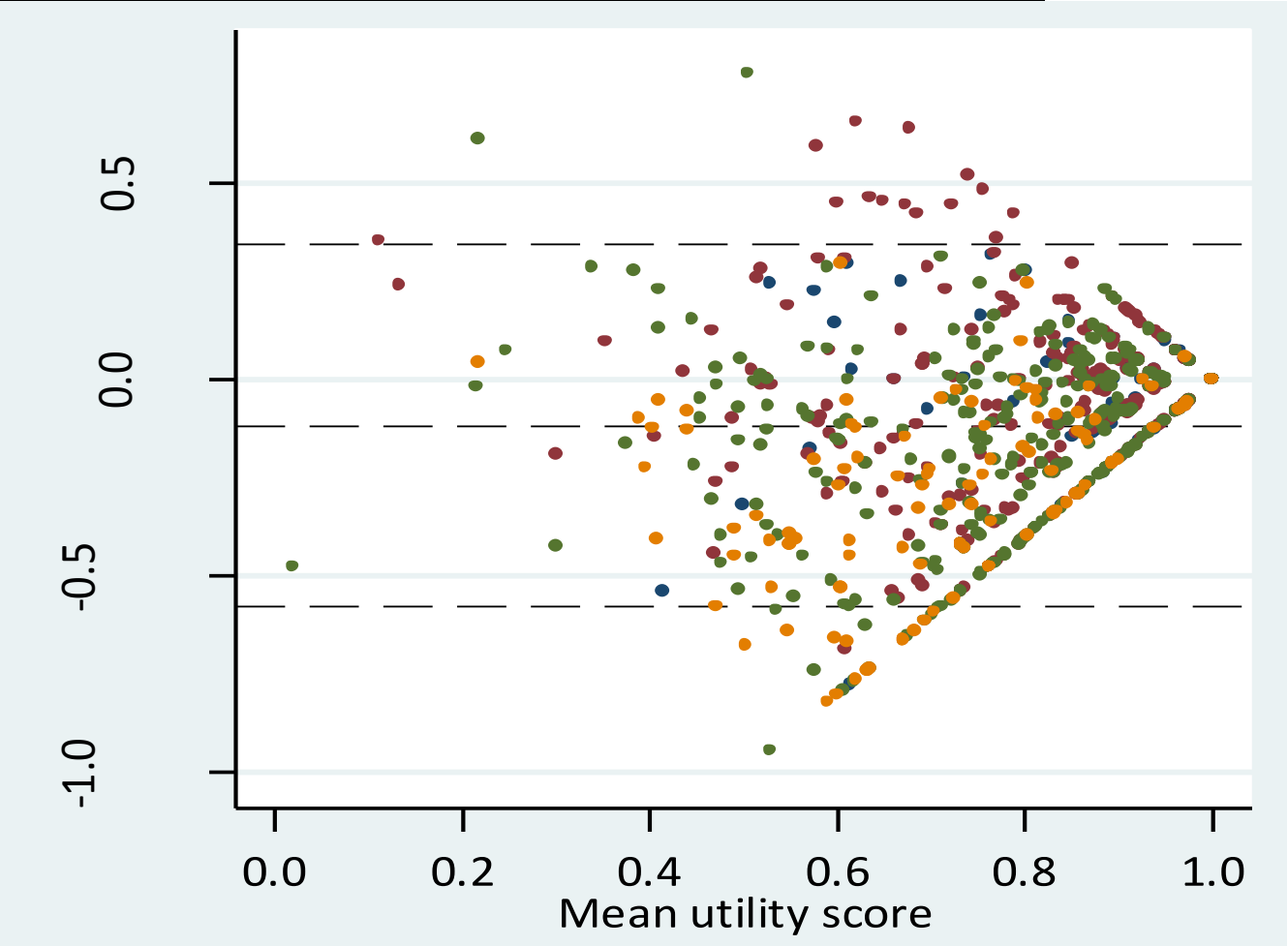


FIGURE A5:

FORMAL-CARER COMPLETED EQ-5D-5L-PROXY COMPARED TO RESIDENT-COMPLETED EQ-5D-5L

	N	rho	Mean difference (95% CI)	LoA
Full sample	894	0.264	-0.118 (-0.133 to -0.103)	-0.579 to 0.343
Subsamples:				
CDR=0	62	0.426	-0.024 (-0.071 to -0.022)	-0.390 to 0.342
CDR=1	341	0.311	-0.037 (-0.059 to -0.015)	-0.445 to 0.371
CDR=2	370	0.259	-0.154 (-0.176 to -0.131)	-0.596 to 0.288
CDR=3	115	0.247	-0.290 (-0.333 to -0.248)	-0.753 to 0.172



Figures A6 to A9:

Results: Comparisons between change in resident-completed and two proxy-completed EQ-5D-5L utility scores over time, by dementia severity

Figure A6:

Change (between two time points) in formal-carer completed EQ-5D-5L score compared to resident-completed EQ-5D-5L score (mean difference = -0.835)

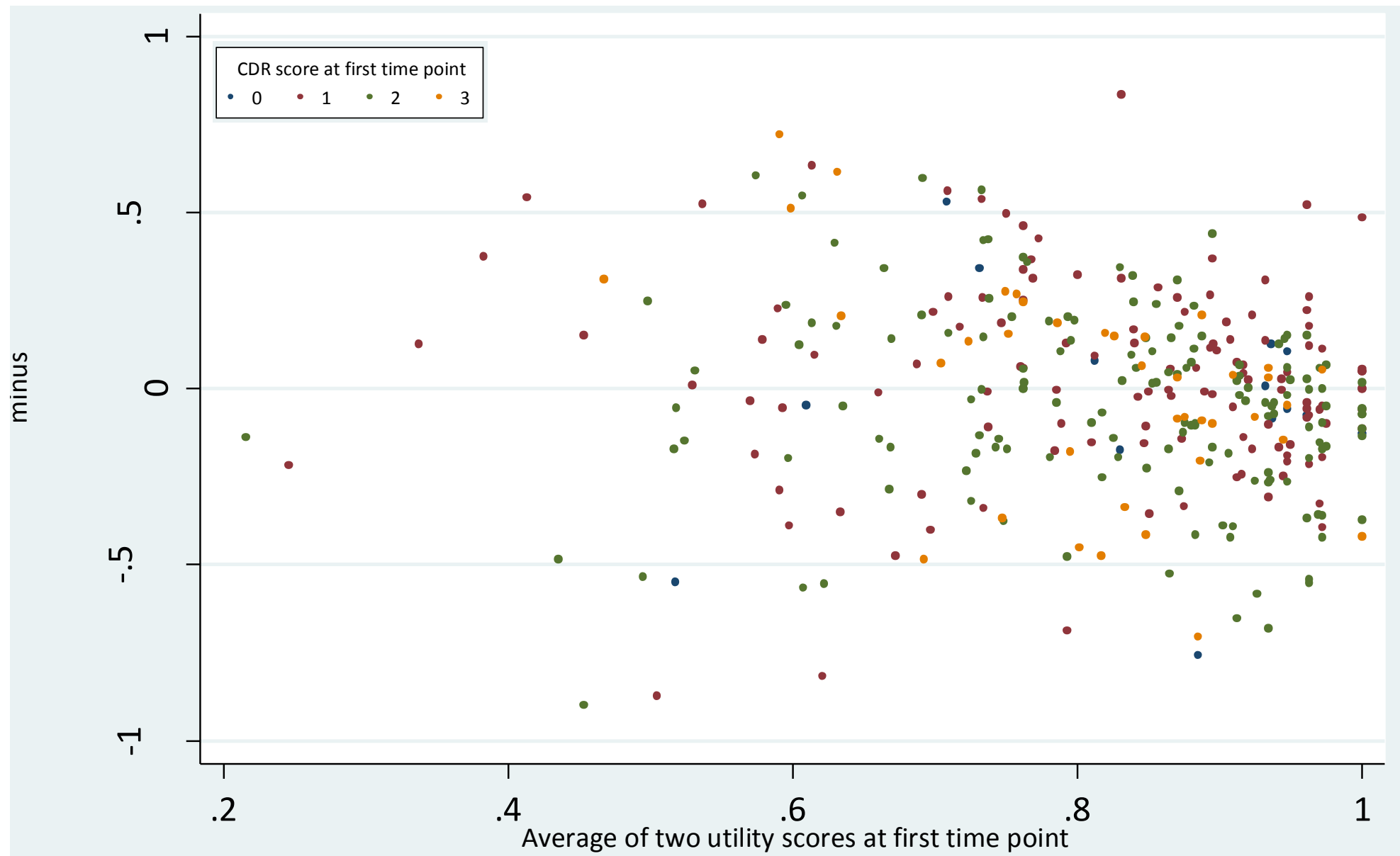


Figure A7:

Change (between two time points) in informal-carer completed EQ-5D-5L score compared to resident-completed EQ-5D-5L score (mean difference = -0.788)

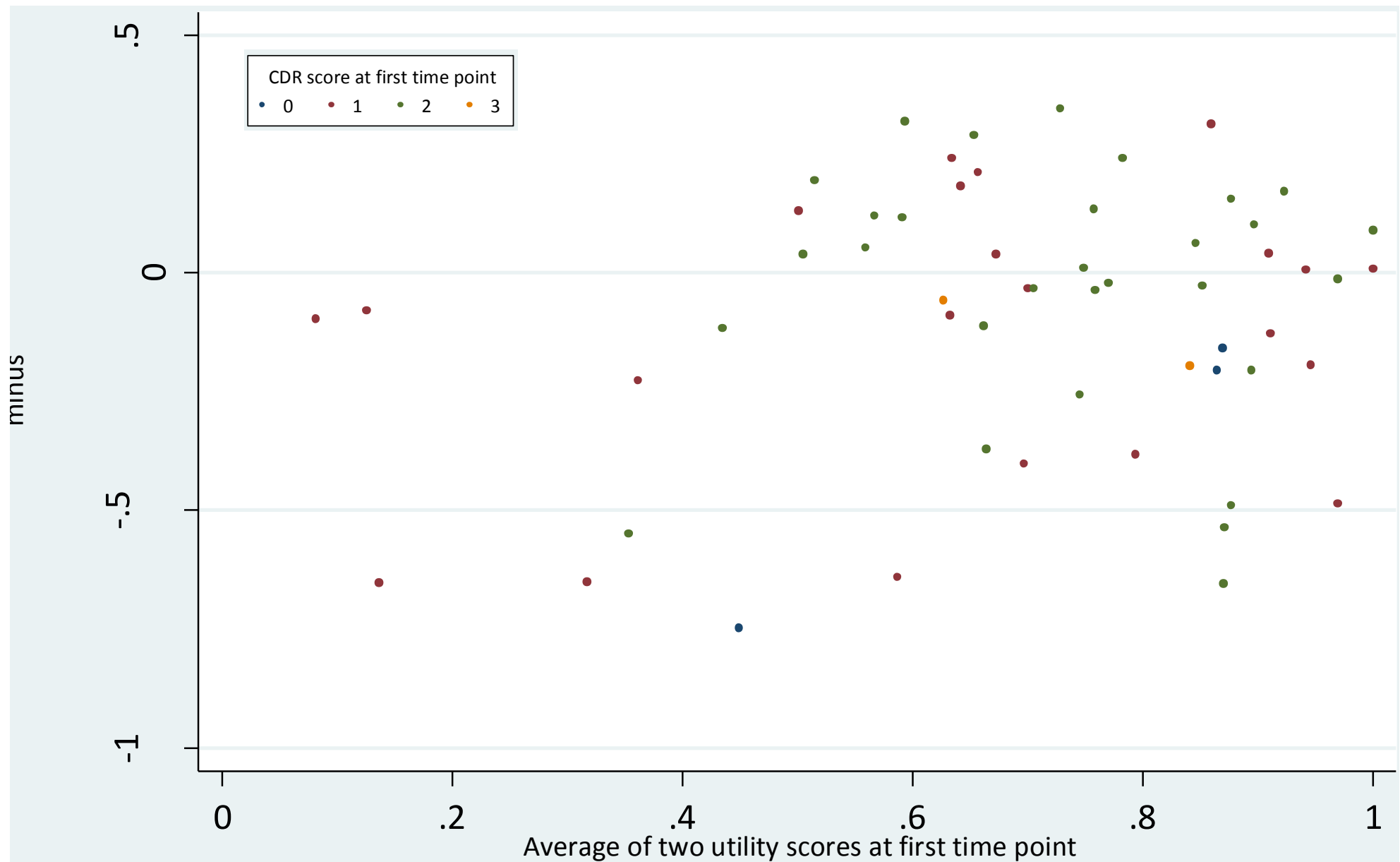


Figure A8:

Change (between two time points) in formal-carer completed DEMQOL-Proxy-U score compared to resident-completed EQ-5D-5L score (mean difference = -0.803)

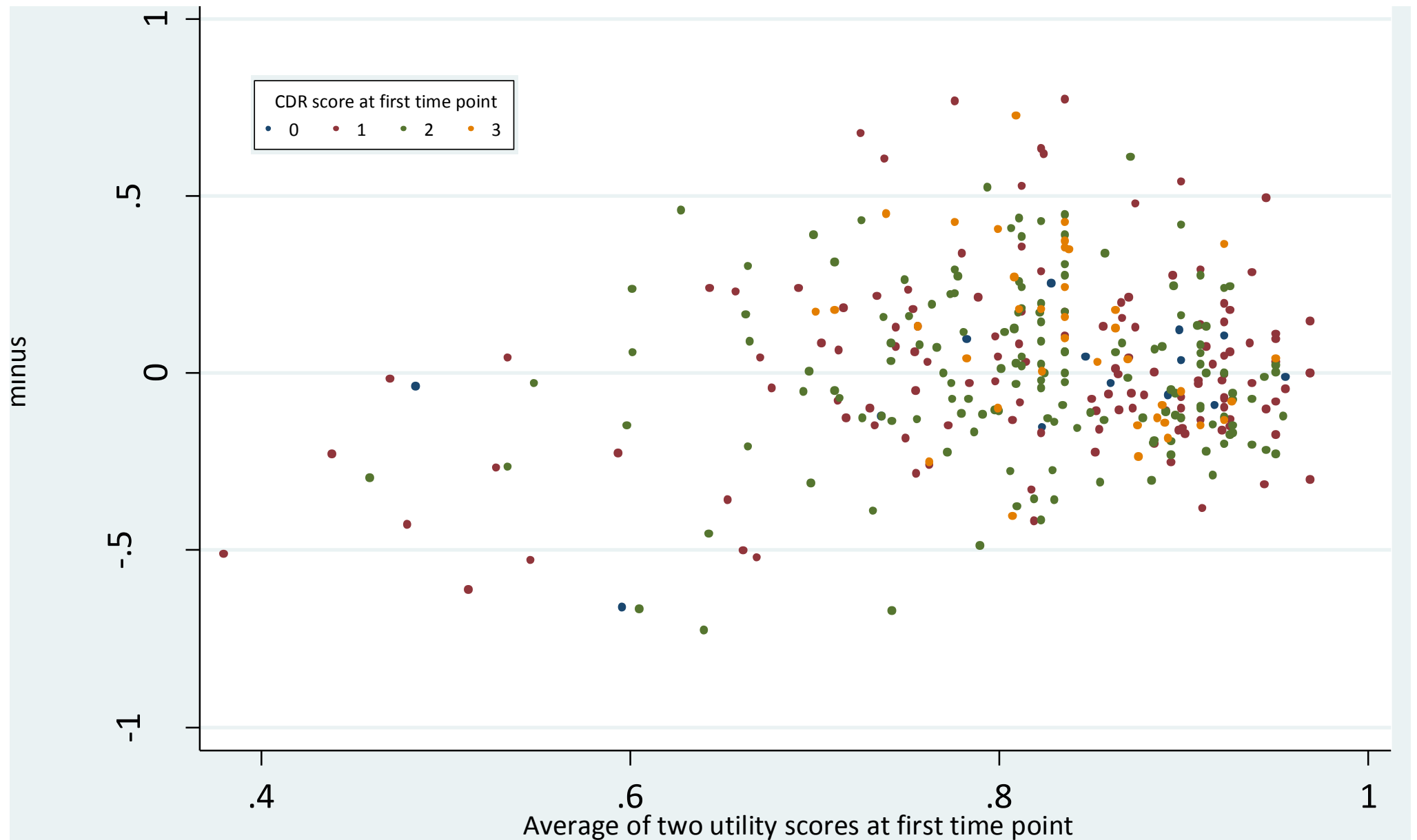


Figure A9:

Change (between two time points) in informal-carer completed DEMQOL-Proxy-U score compared to resident-completed EQ-5D-5L score (mean difference = -0.853)

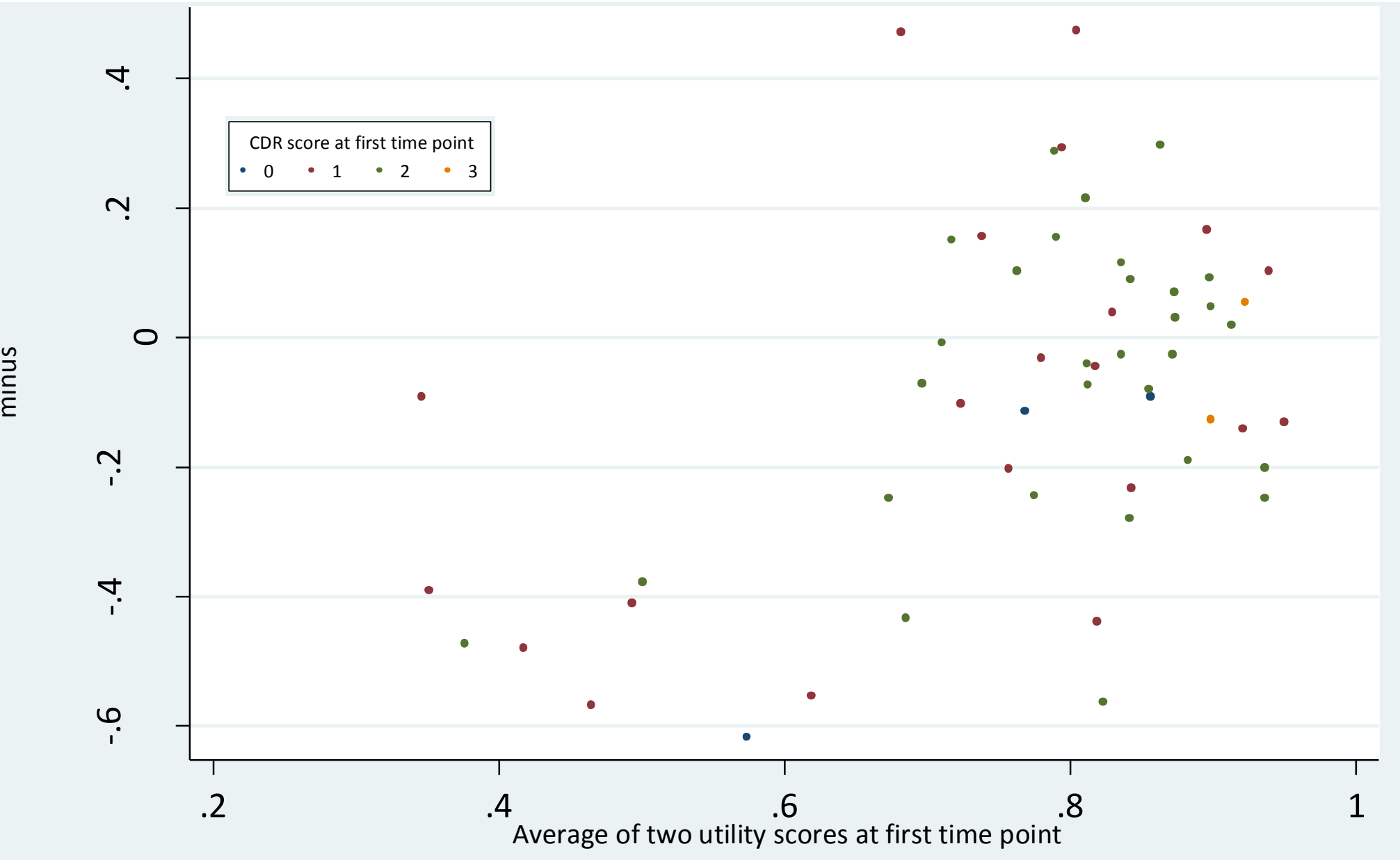




Figure A10: Associations between utility scores and non-utility cognitive and HRQOL measures (criterion validity), excluding incomplete cases (complete cases: N=146, n=92, mean observations per individual=1.604)

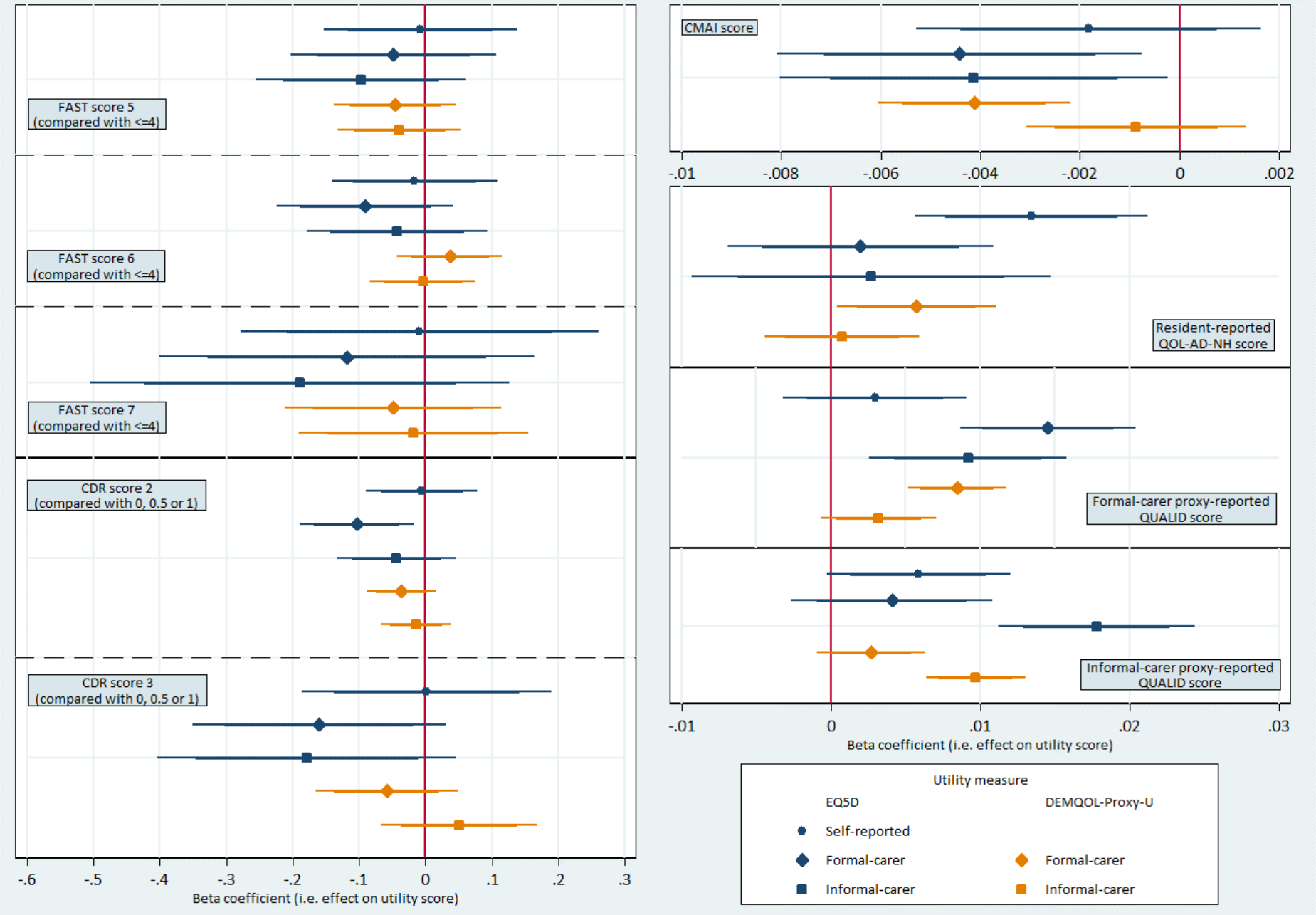
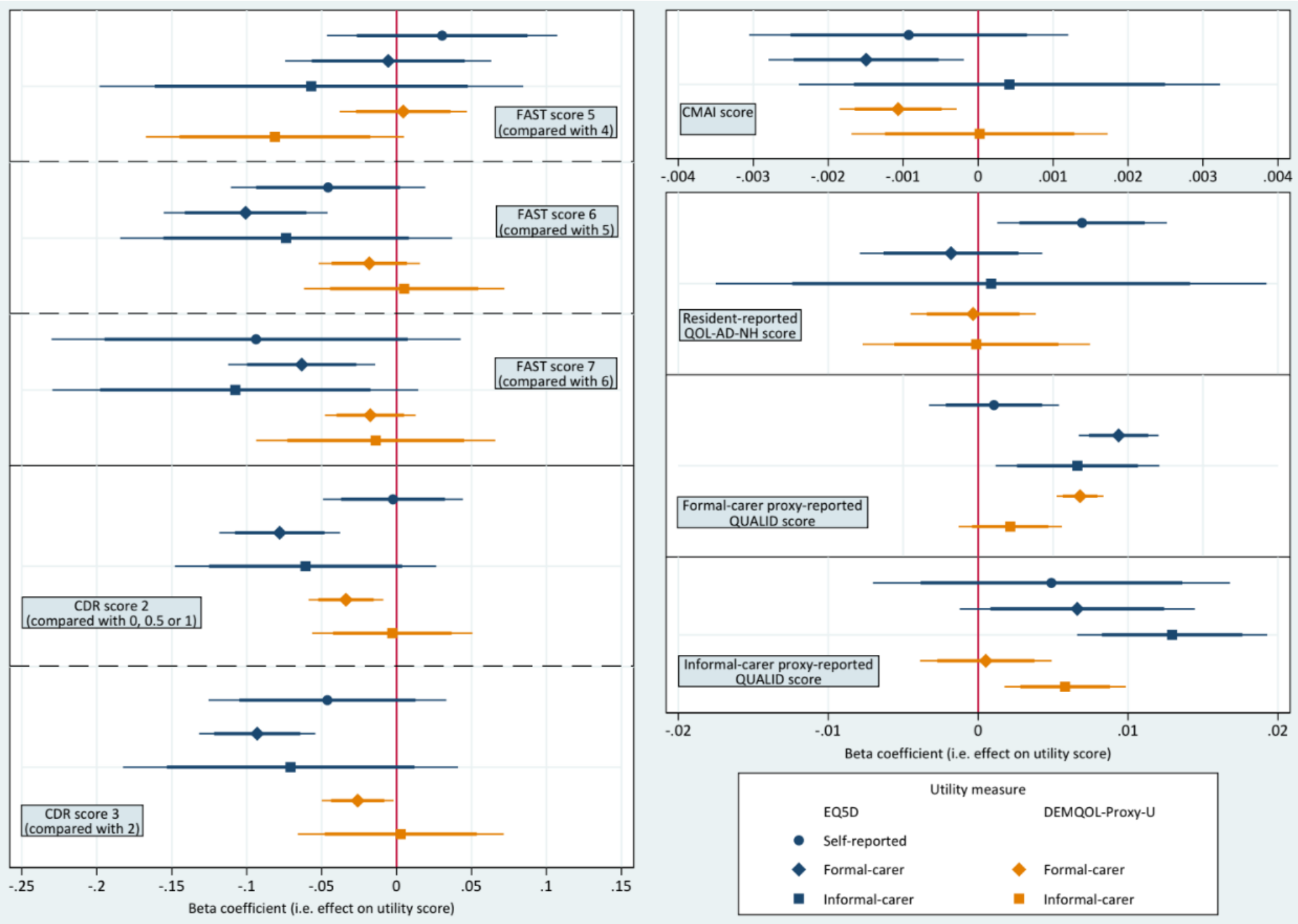


Figure A11: Responsiveness of utility scores to changes in non-utility cognitive and HRQOL measures



**Notes (Figures A10 and A11):**

Figures show the  $\beta_2$  coefficients for model 1 run for each combination of utility (n=5) and non-utility cognitive or HRQOL measure (n=6).

The thicker confidence intervals indicate  $p < 0.05$ . The thinner confidence intervals are after adjustment for multiple hypothesis testing using Bonferroni correction.

The CMAI (range 29-203), QOLAD and QOL-AD-NH (45 point scales) scores are continuous variables.

Figure 10 shows criterion validity using a random effects model, excluding incomplete cases (complete cases: N=146, n=92, mean observations per individual=1.604). The CDR and FAST scores are categorical variables. CDR 0, 0.5 and 1 are in the reference category. FAST  $\leq 4$  is in the reference category.

Figure 11 shows responsiveness using a fixed effects model. The FAST and CDR scores are categorical variables. The  $\beta_2$  coefficients for FAST score 5 is compared here to the  $\beta_2$  coefficient for FAST score 4, coefficients for FAST score 6 are compared to FAST score 5, and FAST score 7 is compared to FAST score 6. Similarly, CDR score 2 is compared to the reference category (i.e. CDR 0, 0.5 and 1), and CDR score 3 is compared to CDR score 2.